

UNIVERSIDAD DE ALICANTE

FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES



GRADO EN ADMINISTRACIÓN Y DIRECCIÓN EMPRESAS

CURSO ACADÉMICO 2019 - 2020

ANÁLISIS DEL PRECIO DE LOS ALOJAMIENTOS AIRBNB DE VALENCIA
APLICANDO MÉTODOS DE REGRESIÓN ESPACIAL

AUTOR: ALEJANDRO BELTRÁN VERDÚ

TUTOR: PEDRO ALBARRÁN PÉREZ

DEPARTAMENTO DE FUNDAMENTOS DEL ANÁLISIS ECONÓMICO

Alicante, Julio 2020

RESUMEN

El objetivo de este trabajo es analizar los precios del mercado de Airbnb en la ciudad española de Valencia para encontrar la forma de ajustar un valor justificado a cada alojamiento, dependiendo tanto de características internas del alojamiento como de características espaciales de la zona en la que esté ubicado. Para ello, se ha ejecutado el método tradicional de regresión lineal múltiple (OLS) y el de regresión ponderada geográficamente (GWR). Los procesos empleados han permitido seleccionar variables que afectan al precio y evaluar los distintos impactos que tienen éstas según la zona o distrito de la ciudad en el que se encuentre la oferta. Además, los resultados han confirmado una mejor actuación del modelo GWR debido a su capacidad de aplicar variación espacial a los efectos que produce el modelo.

PALABRAS CLAVE

Airbnb; precios; Valencia; alojamientos; OLS; GWR

ÍNDICE

1. Introducción	4
2. Revisión bibliográfica	5
3. Objetivos	10
4. Metodología	10
4.1. Recolección e integración de datos.....	10
4.2. Estimación del modelo.....	15
4.2.1. Regresión lineal múltiple (OLS)	15
4.2.2. Regresión ponderada geográficamente (GWR).....	16
5. Resultados	17
5.1. Análisis de los datos de Airbnb Valencia.....	17
5.1.1. Precio de los alojamientos (variable dependiente).....	17
5.2. Modelos de regresión lineal (OLS).....	21
5.3. Modelo de regresión ponderada geográficamente (GWR).....	25
6. Conclusiones	31
7. Bibliografía	34
8. Anexos	37

1 INTRODUCCIÓN

La economía colaborativa es un sistema bastante actual que surge a principios del siglo XXI. Se basa en un sistema de interacción entre dos o más personas con la finalidad de satisfacer una necesidad. En los últimos años ha ido creciendo a un nivel muy considerable debido, en gran parte, al acelerado progreso experimentado por las tecnologías de la información. Esta economía ha tenido un gran auge en diversos sectores como el turismo (Airbnb, HomeAway), transporte (Uber, Cabify, Blablacar) y el sector financiero (Mintos), siendo potenciada por la existencia de modelos de plataformas peer-to-peer (P2P) en el que las interacciones se realizan “entre iguales”, es decir, relaciones que se desarrollan entre dos sujetos que se encuentran en similar posición. También se conoce como consumer-to-consumer (C2C). Además, el hecho de que el servicio se gestione sin recurrir a un sujeto profesional posibilita que se preste a un precio significativamente menor.

El enfoque principal de este trabajo corresponde al mercado de alojamiento turístico y más concretamente a la compañía Airbnb, cuyo nombre viene originado por el acrónimo “Air-bed and breakfast”. Fue creada en 2008 en San Francisco (California) y se dedica a la oferta de alojamiento a particulares y turistas, teniendo la oportunidad de valorarse mutuamente para establecer una calificación de referencia para futuros usuarios. Está considerada como una de las plataformas de economía colaborativa más exitosas en la actualidad (junto a Uber) y, actualmente, dispone de más de 7 millones de alojamientos en más de 220 países y 100.000 ciudades del mundo, con un total de más de 750 millones de huéspedes a lo largo de su trayectoria, según la propia página web de Airbnb. España es el cuarto país del mundo (por detrás de Estados Unidos, Francia e Italia) con más alojamientos, ascendiendo a un total de 245.000.

Este estudio pretende realizar un análisis enfocado al precio de los alojamientos Airbnb de Valencia, la tercera ciudad (por detrás de Barcelona y Madrid) con más oferta del país, tratando de explicar cómo se fijan los precios de esta compañía. Teniendo en cuenta que la compañía no está sujeta a ninguna política reguladora de precios y los anfitriones son libres de ofertar su propiedad al valor que creen conveniente, el modelo se ha estimado en el contexto del método de precios hedónicos. Este método analiza los atributos de un bien, generalmente estimados en un modelo de regresión, donde el precio es la variable dependiente y emplea los coeficientes de cada variable independiente añadida al modelo para estimar su valor. De esta forma, se ha decidido estudiar una serie

de indicadores que explican qué precio se debería establecer (aproximadamente) como vendedor y, a la vez, qué precio resulta justo para el consumidor del servicio en el mercado local de la ciudad. No solo se consideran aspectos internos del alojamiento, sino también factores espaciales que pueden llegar a provocar variaciones en el precio dependiendo de su localización. Para ello, se han recolectado datos espaciales de la ciudad y se han aplicado a modelos de regresión lineal con el objetivo de ejecutar un posterior modelo de regresión ponderada geográficamente mediante el uso del software RStudio como Sistema de Información Geográfica (SIG).

El resto del trabajo está organizado de la siguiente manera. En la sección 2 se han evaluado estudios anteriores acerca de cuestiones y procesos similares al tema que se va a tratar. La sección 3 plantea los objetivos que se tratan y desarrollan a medida que avanza el trabajo. La sección 4 detalla la metodología que especifica los procesos de recolección de datos y modelado estadístico. En la sección 5 se desarrollan los resultados fruto de la aplicación de los métodos especificados en la metodología. En la sección 6 se argumentan las conclusiones obtenidas a partir del análisis de resultados obtenidos. Por último, la sección 7 hace referencia a las fuentes bibliográficas y la 8 muestra los anexos de algunas tablas y figuras que no son de especial interés, pero se pueden consultar para entender mejor algunos procesos.

2 REVISIÓN BIBLIOGRÁFICA

España tiene unos elevados números en la presencia de Airbnb principalmente por el hecho de ser el segundo país (por detrás de Francia) con más afluencia turística del mundo, recibiendo una cantidad de 83,7 millones de turistas internacionales en 2019 y suponiendo un incremento del 1,1% respecto al año anterior (INE).

Estudios anteriores han afirmado que la aparición de Airbnb en las ciudades ha encarecido el precio de los alquileres, ya que la oferta de éstos se ha visto disminuida porque muchos arrendadores han optado por utilizar su propiedad a modo de alojamiento turístico en plataformas como Airbnb. Guttentag (2013) estudió Airbnb como una innovación disruptiva en el sector de alojamiento. El autor lo vincula con el término “disruptivo” por su transformación de mercado con un conjunto distinto de beneficios, enfocado principalmente en ser más barato, más conveniente o más simple.

Algunos autores han realizado estudios sobre los efectos que puede tener la oferta Airbnb sobre el mercado del alquiler y compraventa de vivienda. Barron, Kung y Proserpio (2019) analizan el impacto de Airbnb en el mercado inmobiliario de las 100 áreas metropolitanas más grandes de los Estados Unidos. Sus conclusiones indican que la subida de un 1% de la actividad de la compañía, provoca un incremento de los precios de alquiler de la zona en un 0,018% y los de compra en un 0,026%, reflejándose de esta forma en un aumento medio mensual del alquiler de 9\$ y unos 1800\$ más en el precio de compra de la vivienda.

Por otra parte, Koster, van Ommeren y Volkhausen (2018) evaluaron el impacto de la inhabilitación de Airbnb en áreas específicas de Los Ángeles. Esta situación les permitió aplicar técnicas econométricas que comparan ubicaciones muy cercanas separadas únicamente por una frontera administrativa para evaluar los efectos de la prohibición. Los autores concluyeron que la prohibición redujo la actividad Airbnb en un 50% y que, en consecuencia, los precios de alquiler cayeron un 3%. A través de estos resultados, se puede comprobar el impacto directo de Airbnb en el mercado del alquiler.

Según un estudio del Instituto de Economía de Barcelona (IEB), la penetración del mercado Airbnb en la ciudad con mayor presencia de este tipo de alojamientos de España ha tenido un impacto de subida del 19% en los precios de compraventa y un 7% en el mercado de alquiler entre 2012 y 2016. Se calcula que por cada 54 anuncios activos de Airbnb en un barrio, los alquileres se encarecen un 1,9%, mientras que el precio de compraventas repunta más del doble, un 5,3%. En 2015, Airbnb representaba el 6,84% de las viviendas en alquiler de Barcelona, siendo este porcentaje una cantidad bastante significativa teniendo en cuenta que otras ciudades como París, Nueva York o Los Ángeles se encuentran por debajo del 5%. Esta decantación por este tipo de negocio se debe principalmente al aumento sustancial de la renta de los arrendadores, ya que el servicio de alquiler que ofertan en la plataforma reporta en apenas 10 días el equivalente a un mes de renta, triplicando el beneficio mensual en el caso de plena ocupación.

En 2018, la Comunidad Valenciana recibió un total de 9,2 millones de turistas, aumentando así un 3,2% con respecto al año anterior. Según las Estadísticas de Turismo de Valencia, el turismo en la ciudad mantuvo la evolución positiva que llevaba teniendo cada año con un aumento de viajeros del 2,8% (2.064.094). Dentro de estas cifras, se puede destacar el turismo internacional que se sitúa en el 65% del total, sobre todo franceses y británicos, con un aumento del 69% y 68%, respectivamente, con relación al año al año anterior.

La ocupación media en la ciudad creció un 16% hasta situarse en un 74%. El total de hoteles que hay en la ciudad es de 164, ofertando un total de 18.986 plazas disponibles. Las pernoctaciones hoteleras registraron un incremento interanual del 3,9%. Estas cifras contestan a una de las preguntas más habituales que se plantean: ¿los alojamientos Airbnb suponen una fuerte competencia para el sector hotelero o un complemento a su oferta? La respuesta es clara: hasta ahora han sido una oferta complementaria demostrada por la constante tendencia de crecimiento que las pernoctaciones hoteleras han mantenido.

Estas afirmaciones se ven respaldadas por un estudio realizado por la Escuela Hotelera de Lausanne (EHL) en 2018, en el que analizó el impacto de Airbnb en su lugar de origen (San Francisco) donde alegaba que el volumen total no afecta al RevPar (Revenue per Available Room) de los hoteles, sino que lo complementa. Sin embargo, sí que tiene un impacto en el precio medio de los hoteles, ya que cuanto mayor sean las tarifas de los alojamientos ofertados en la web, mayor será el RevPar de los hoteles. El cálculo del RevPar se realiza dividiendo los ingresos por habitación entre habitaciones disponibles.

Los resultados obtenidos brindan “evidencia empírica” de que Airbnb interrumpirá el negocio hotelero con el tiempo; cuantos más usuarios de la plataforma estén satisfechos con su experiencia, es más probable que disminuya la demanda de habitaciones de hotel. Los hoteles serán preferibles para un target que busque una experiencia más cómoda o un alojamiento con posibilidad de disfrutar de una serie de actividades dinámicas dentro del propio recinto. Los turistas que busquen un turismo más enfocado a disfrutar del exterior y a utilizar el alojamiento tan solo para las necesidades básicas, preferirán una opción más barata y que permita una mayor libertad en su propio espacio privado, sin depender de ciertos horarios establecidos.

Según los datos de Hosbec (2018), una asociación empresarial hotelera, en el año 2017 se comercializaron en Valencia una media de 5.495 apartamentos y habitaciones, oferta que puede hospedar alrededor de 22.000 turistas, mientras que la oferta hotelera es de 16.200 plazas, señalando así que Airbnb supera la oferta hotelera al menos en un 35%. Considerando el informe elaborado por su departamento de Big Data, la tarifa media por noche se sitúa en 60 y los ingresos medios que obtienen por operación se sitúan sobre los 802 euros. Esta cifra asciende a 80 euros por noche en temporada alta, siendo ésta en marzo (debido a la conocida fiesta valenciana “Fallas”) y los meses de verano. La tasa de ocupación de los apartamentos fue del 60%. Además, mediante una estadística realizada, indica que existe un 72% de anfitriones que poseen más de un alojamiento en la

plataforma, significando esto que supone una actividad económica habitual para muchos usuarios.

A través de la elaboración de un gráfico con la evolución mensual en el año 2019 (INE), se puede apreciar cómo existe un aumento significativo de viajeros y pernoctaciones en el mes de marzo, siguiendo una tendencia incremental hasta el mes de septiembre, donde Valencia recibe su máximo de turistas.

De acuerdo con Zervas, Proserpio y Byers (2014), a diferencia de los hoteles, Airbnb puede ampliar la oferta en cualquier lugar con casas y apartamentos ya existentes, mientras que los hoteles deben construirse siguiendo unos requerimientos regionales dependientes de cada zona. De esta manera, en comparación con los hoteles, Airbnb puede extender más fácilmente sus servicios al centro histórico, y el hotel no solo necesita todo el edificio para establecer su propia ubicación, sino que también requiere el permiso de las autoridades.

Los hoteles se han tenido que ver obligados a “fidelizar el cliente con su experiencia en el hotel”, según el Instituto Tecnológico (ITH). Este hecho se confirma por un gran problema que afrontan, quizás el problema principal, que es el ahorro que supone a los clientes reservar un alojamiento Airbnb en lugar de reservar un hotel. En algunas ciudades, el precio de este tipo de alojamiento se reduce a más de la mitad del que los turistas tienen que pagar en un hotel. La firma de capital riesgo Kleiner Perkins hizo público en 2018 una comparativa entre los precios de Airbnb y los hoteles de varias ciudades del mundo. En el caso de Tokio, la diferencia de precios entre un alojamiento hotelero y la firma estadounidense es la más destacada, suponiendo un 57,73% más caro. Otras ciudades como Nueva York, Moscú o Sidney experimentaban una diferencia del 38,89%, 44,92% y 20,42%, respectivamente. Está comprobado que esta diferencia de precios es una competencia muy difícil de afrontar para el sector hotelero, motivo por el que se movilizaron en contra de la “no regulación” de precios de este tipo de alojamientos.

Las gráficas de variación de la ocupación de los hoteles indican que su demanda ha seguido creciendo a pesar del fenómeno Airbnb. En cambio, los alojamientos que más perjudicados se han visto han sido los hostales y las pensiones debido a su gran similitud de características.

Valencia, junto a otras ciudades (Barcelona, Amsterdam, Berlín, Munich, Burdeos, Bruselas, Cracovia, París y Viena), reclamaron en 2019 el fin de la “carta blanca” de la que dispone Airbnb. Esto se debe a su efecto sobre el precio de los alquileres de las grandes ciudades y a la expulsión de los vecinos de barrios históricos debido a las

externalidades negativas que provoca. Estas externalidades negativas son, en mayor parte, actividades y horarios que mantienen los turistas que se hospedan temporalmente. El análisis de García-Palomares, Gutiérrez y Mínguez (2015) confirmó que estas áreas centrales con las zonas de mayor concentración de turistas conllevaban a problemas de gentrificación. Además, muestran que la distribución de los Airbnb en Barcelona tiene un claro patrón centro-periferia. Sus alojamientos tienden a estar concentrados en el centro de la ciudad y su distribución es simple y regular, desde los clusters de valores altos-altos (H-H) en el centro a los clusters de valores bajos-bajos (L-L) en la periferia.

Fernandes (2018) estudia los precios del mercado Airbnb en Lisboa. En su análisis aborda asuntos que se llevarán a cabo en este trabajo, tratando de hallar un modelo GWR que explique las variaciones espaciales de los datos según su localización. Se comprueban características que justifican el uso de este modelo, como la existencia de autocorrelación espacial en el precio de los alojamientos. Mediante la elaboración de mapas, muestra visualmente el impacto de los coeficientes locales de las variables explicativas del modelo en las distintas zonas de la ciudad, pudiendo así identificar los distritos en los que los coeficientes tienen más o menos peso.

Lansley y Cheshire (2016) realizaron un estudio del distrito de Camden de la ciudad de Londres. El principal objetivo en el que se basa su análisis es ejecutar una predicción del porcentaje de gente con altas cualificaciones atendiendo a la utilización de ciertas variables explicativas que le permiten realizar un modelo lineal en primera instancia. Tras comprobar sus resultados, se observa que la aplicación de un análisis espacial de la zona podría proporcionar mejores resultados. De esta forma, amplifican el proceso elaborando un análisis de regresión ponderada geográficamente, donde los coeficientes obtenidos del modelo pueden variar a través del espacio. Este modelo de regresión complementa al análisis de regresión lineal tradicional, utilizando las coordenadas de cada observación de la zona de estudio para la aplicación del método. El proceso elaborado por estos autores ha servido como guía orientativa en el desarrollo del método de este estudio aplicado a los precios.

3 OBJETIVOS

Existen varias cuestiones que se plantean con relación a la actuación y distribución de los precios de los alojamientos.

¿Hay alguna dependencia espacial entre los alojamientos Airbnb? La autocorrelación de una variable específica permite analizar si las diferentes observaciones guardan relación con un número determinado de entidades consideradas vecinas por cercanía.

¿Se pueden observar lugares donde hay concentración de valores altos o valores bajos de los precios? A través de un análisis hot spot se puede comprobar visualmente en qué zonas del mapa de Valencia existe mayor concentración de valores altos o bajos de una variable específica.

¿Qué variables disponibles influyen en la variación del precio? Mediante el análisis de regresión lineal, se puede observar el nivel de significatividad que los atributos añadidos al modelo tienen sobre la variable dependiente y el peso de los coeficientes de estos predictores.

¿Tienen diferente impacto estas variables según la ubicación de los alojamientos? El modelo GWR permite que los efectos de los resultados del modelo de regresión tengan variación espacial.

Estas son algunas de las preguntas que se desarrollarán con más profundidad en el desarrollo del método.

La meta general conlleva a la realización de varios procesos que se deben llevar a cabo:

- Creación de datos con información espacial;
- Investigar patrones de autocorrelación espacial en los datos;
- Estimar modelos de regresión lineal (OLS) y hallar el más eficaz;
- Estimar un modelo de regresión ponderada geográficamente (GWR);
- Comparar la actuación de los modelos.

4 METODOLOGÍA

4.1 Recolección e integración de datos

Esta investigación está principalmente basada en un análisis de datos de los alojamientos Airbnb en la ciudad de Valencia. La plataforma no proporciona sus datos oficiales, pero existen otras plataformas alternativas en la que investigadores se encargan

de la recogida de sus datos a través del web scrapping. Algunas de ellas son Inside Airbnb y Tom Slee, que contienen bases de datos en distintos formatos de una selección de grandes ciudades del mundo. En este caso, Tom Slee dispone de los datos de los siguientes territorios españoles: Girona, Granada, Madrid, Asturias y Barcelona. Al no tener los datos de la ciudad escogida como objeto de estudio, se ha optado por extraerlos de Inside Airbnb.

Los datos obtenidos están en formato de texto delimitado por comas (.csv) y contienen una gran cantidad de variables. Corresponden al mes de febrero del año 2019. Para el tratamiento de los datos se ha empleado RStudio, un entorno desarrollado integrado para el lenguaje de programación R que permitirá importar, limpiar y analizar datos. El número de variables ha sido seleccionado a partir de los datos originales, escogiendo aquellas que pueden contener un mayor valor descriptivo en la elaboración de los modelos.

Considerando que el objetivo principal de estudio es analizar geográficamente el comportamiento espacial de los precios de Airbnb, se ha necesitado obtener datos espaciales del área de Valencia. A través del catálogo del Portal de Transparencia y Datos Abiertos del Ayuntamiento de Valencia, se pueden conseguir datos sobre transporte, turismo, salud y urbanismo e infraestructuras, entre otros.

Para combinar con los datos, se ha optado por recolectar de los datos abiertos de Valencia la siguiente información:

- División administrativa de los distritos municipales de Valencia;
- Límites espaciales de la ciudad de Valencia;
- Coordenadas espaciales de los equipamientos municipales de Valencia.

Los datos que contienen la información de los límites territoriales y equipamientos municipales corresponden a archivos shapefile (.shp). Un shapefile es un tipo de formato representado por puntos, líneas o polígonos donde sus entidades almacenan atributos e información geográfica. En este caso, el mapa de Valencia se corresponde con un polígono que representa los límites territoriales de la ciudad dividida en sus 19 distritos y sus equipamientos municipales se observan en forma de puntos. Para trabajar con ello, se han unido los datos de Airbnb a los dos shapefile a través de las coordenadas de cada alojamiento para trabajar con las relaciones entre los elementos de los distintos archivos de datos.

Los datos Airbnb recolectados en Inside Airbnb hacen referencia al mes de febrero de 2019. Tras la selección de sus variables más descriptivas, quedan un total de 25 (tanto

cuantitativas como cualitativas). El total de observaciones que contiene asciende a 6303 alojamientos tras limpiar aquellos que aparecían en áreas fuera de los límites territoriales por fallos de proyección. En la Tabla 1 se detallan las variables seleccionadas que se utilizarán en los modelos con sus respectivas descripciones debido a contener información de los atributos internos del alojamiento que tienen relación directa con sus dimensiones. Más habitaciones o más baños generalmente significarán una mayor capacidad de acogida del alojamiento y, por lo tanto, mayor espacio. La longitud y la latitud juegan un papel fundamental en el análisis de los alojamientos en el contexto de los datos de información geográfica importados.

VARIABLE	DESCRIPCIÓN
name	nombre del alojamiento
host_id	número de identificación del anfitrión
host_listings_count	número de alojamientos que tiene el anfitrión
neighbourhood_cleansed	barrio en el que se encuentra el alojamiento
neighbourhood_group_cleansed	distrito en el que se encuentra el alojamiento
room_type	tipo de habitación (privada, compartida o alojamiento entero)
accommodates	capacidad máxima de huéspedes en el alojamiento
bathrooms	número de baños del alojamiento
bedrooms	número de habitaciones del alojamiento
final_price	precio del alojamiento (en \$)
longitude	longitud del alojamiento en formato decimal
latitude	latitud del alojamiento en formato decimal

Tabla 1 – Variables originales

Las variables contienen información útil para entender mejor el mercado local de Airbnb en Valencia. Sin embargo, esta información no es suficiente para analizar de forma precisa el mercado, por lo que se necesita información adicional.

Quattrone et al. (2016) introducen en su análisis de Airbnb en Londres el impacto que tiene la distancia al centro de la ciudad sobre la concentración de alojamientos. Zhang et al. (2017) señalan en su estudio sobre el precio Airbnb en Metro Nashville (Davidson County), Tennessee, el impacto que tienen la distancia al centro y atracciones locales, entre otros, sobre el precio final de los alojamientos Airbnb, aunque en su caso incluía la variable de valoración de los alojamientos. En este estudio no se ha incluido ninguna variable referente a la opinión de los usuarios debido a que la introducción de un nuevo alojamiento en la plataforma no podría basarse en ninguna valoración porque la fijación del precio es un paso previo a la opinión del primer consumidor.

Considerando la eficacia de los resultados en anteriores estudios, se han creado una serie de variables que se detallan en la Tabla 2. Los factores considerados han sido el nivel de demanda del conjunto de equipamientos municipales disponibles en los datos y distancias lineales entre cada alojamiento y lugares enfocados a un turismo específico y generalizado, como el centro de la ciudad y la playa.

VARIABLE	DESCRIPCIÓN
VHigh	nº de equipamientos municipales cercanos de demanda muy alta (<1000m)
High	nº de equipamientos municipales cercanos de demanda alta (<1000m)
Medium	nº de equipamientos municipales cercanos de demanda media (<1000m)
MediumLow	nº de equipamientos municipales cercanos de demanda media-baja (<1000m)
distance_center	distancia (en metros) de cada alojamiento al centro de la ciudad
beach_id	número de identificación de la playa más cercana
distance_beach	distancia (en metros) de cada alojamiento a la playa más cercana

Tabla 2 – Variables creadas

Entre las variables creadas se puede observar la distinción de 4 variables referentes al número de equipamientos municipales en un radio de 1000 metros a cada apartamento según su nivel de demanda (muy alta, alta, media, media-baja). El nivel de demanda se ha escogido a través de un análisis de las opiniones de los turistas en TripAdvisor. El criterio de puntuación se ha basado en el número de opiniones y valoración de cada lugar de interés (Tabla A1 – Anexo). También se ha creado la variable de distancia al centro, representado por la Plaza del Ayuntamiento, expresada en metros a cada uno de los alojamientos ofertados. En este estudio, se introduce una nueva variable de influencia que se puede considerar significativa en el precio debido a la gran cantidad de turismo de sol y playa que la ciudad también recibe. Esta variable es la distancia a la playa, donde se han recogido 6 localizaciones referentes a distintos puntos de las playas de Valencia, entre las que destaca principalmente la Playa de la Malvarrosa.

Estas variables espaciales se han creado a través de un software destinado al tratamiento de información geográfica. Para ello, se han combinado los shapefile de los límites territoriales de Valencia y los equipamientos municipales más importantes, junto con los puntos de las observaciones de Airbnb. El programa utilizado es QGIS, con el que a través de las coordenadas de los elementos mencionados y características asociadas a ellos, se puede tratar la información de una forma más gráfica. Es una de las grandes ventajas de los SIG, analizar grandes cantidades de datos que estén vinculados a una referencia espacial.

El método utilizado para la creación de las variables finales ha sido la matriz de distancia, donde se calcula la distancia entre los puntos de entrada (alojamientos Airbnb) y los de salida (equipamientos municipales seleccionados o ubicaciones de playas). En el caso de la distancia a la playa, se ha basado en el criterio de medir la distancia entre cada alojamiento y el punto de localización de playa más cercana. Estos datos obtenidos a partir de la matriz de distancia se han exportado de nuevo a RStudio para su aplicación en los modelos.

Teniendo en cuenta las coordenadas geográficas de los alojamientos Airbnb, se puede combinar con el shapefile de los límites territoriales de la ciudad y ver la distribución espacial de los alojamientos en Valencia.

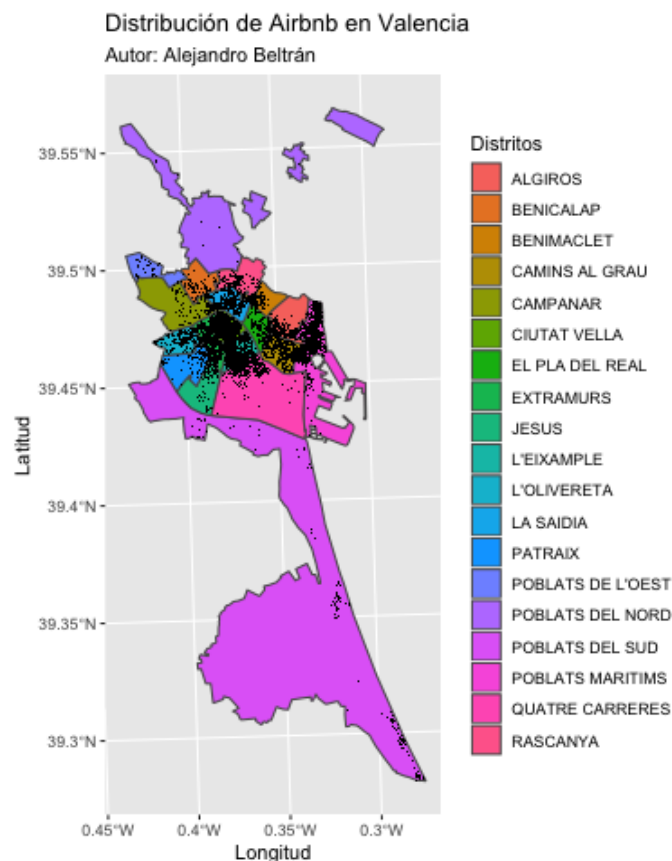


Figura 1 – Distribución de los alojamientos Airbnb en Valencia

En la Figura 1 se puede apreciar la distribución de los alojamientos Airbnb (puntos negros) dentro de la ciudad, con una clara concentración en las zonas del centro de la ciudad y la costa. El mapa se ha realizado mediante la unión del shapefile de la división administrativa del territorio valenciano y los datos Airbnb.

4.2 Estimación del modelo

Existen varias investigaciones que han utilizado la regresión ponderada geográficamente para estudiar diferentes situaciones y materias. Wang et al. (2015) aplican este modelo en el análisis de las diferencias espaciales de la huella de carbono en la China provenzal de 2010.

La regresión espacial ha sido utilizada también por muchos investigadores para estudiar patrones espaciales en varias ciudades. Zhang et al. (2017) emplean este método para calcular el precio de los alojamientos Airbnb. En su análisis utiliza métodos de regresión lineal como OLS, pero alegan que estos métodos no tienen en cuenta la localización de las observaciones en el análisis, ya que solo busca relaciones entre la variable dependiente y sus correspondientes variables explicativas, omitiendo así la heterogeneidad espacial en las relaciones entre alojamientos.

A pesar de que está claramente probado que la regresión ponderada geográficamente proporciona unos mejores resultados en este tipo de análisis, el primer paso a realizar es encontrar el modelo OLS que mejor predice y a continuación aplicar las mismas variables explicativas en la estimación del modelo espacial. El desarrollo de los modelos y la interpretación de los resultados se ha realizado en RStudio.

4.2.1 Regresión lineal múltiple (OLS)

En los modelos de regresión lineal múltiple, el método más popular es el de Mínimos Cuadrados Ordinarios (OLS) debido a sus propiedades descriptivas y facilidad de cálculo. En él, la variable dependiente se estima mediante una serie de variables explicativas. Se expresa de la siguiente forma:

$$y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + e_i$$

donde

- y_i – observación número i de la variable dependiente;
- α (intersección) – término independiente. Valor predicho de y cuando todas las variables explicativas son igual a 0;
- x_i – variables explicativas;

- $\beta = (\beta_1, \beta_2, \dots, \beta_n)$ – coeficientes de los predictores estimados en el modelo;
- e_i – error de observación debido a variables no controladas.

4.2.2 Regresión ponderada geográficamente (GWR)

La regresión ponderada geográficamente es una forma local de regresión lineal que se utiliza para modelar las relaciones entre la variable dependiente y explicativas que varían espacialmente. De este modo, se ejecuta una ecuación de regresión para cada observación de la muestra.

Mediante el uso de GWR, se construye una ecuación de regresión distinta para cada observación del conjunto de datos al incorporar las variables dependiente y explicativas de las entidades vecinas que caen dentro de un ancho de banda que se halla mediante el cálculo de un kernel (núcleo) que puede ser tanto fijo como adaptativo. Ambos tipos de kernel utilizan una función gaussiana, con la diferencia de que el primero calcula un ancho de banda fijo que puede resultar ineficiente en zonas donde los datos son escasos, mientras que el segundo calcula el número óptimo de observaciones vecinas para cada entidad. En este estudio se ha optado por emplear un tipo de kernel adaptativo.

El modelo GWR se expresa de la siguiente forma:

$$y_i = \alpha(u_i, v_i) + \beta_1(u_i, v_i)x_{i1} + \beta_2(u_i, v_i)x_{i2} + \dots + \beta_n(u_i, v_i)x_{in} + e_i$$

donde

- y_i – variable dependiente en la localización i ;
- α (intersección) – término independiente. Valor predicho de y en la localización i cuando todas las variables explicativas son igual a 0;
- x_{in} – valor de la variable explicativa número p en la localización i ;
- $\beta = (\beta_1, \beta_2, \dots, \beta_n)$ – coeficientes de los predictores estimados en el modelo;
- (u_i, v_i) – coordenadas de la localización i ;
- e_i – error de observación debido a variables no controladas en la localización i .

5 RESULTADOS

5.1 Análisis de los datos de Airbnb Valencia

El principal objetivo de este apartado es el de ofrecer una vista más profunda sobre los datos que se han utilizado para el análisis. En los datos obtenidos inicialmente, se ha tenido que hacer una selección de las variables más útiles y resultados descriptivos de sus características como los valores atípicos que conllevan a una alteración de los resultados óptimos del modelo. Se ha realizado un gráfico de correlación (Figura A1 - Anexo) entre la variable dependiente (precio de los Airbnb) y las variables explicativas y entre las propias variables explicativas de los modelos, utilizando el método Pearson. Algunas variables potencialmente explicativas presentan altas correlaciones entre ellas que pueden conllevar a problemas en la elaboración de los modelos de regresión.

En los resultados de la matriz se espera tener una alta correlación entre la variable dependiente y sus variables explicativas, mientras que entre las variables explicativas se espera obtener una baja correlación para evitar problemas de multicolinealidad en el modelo, donde simplemente se eliminará una de las variables que la provocan.

5.1.1 Precio de los alojamientos (variable dependiente)

El precio toma un papel importante en el ámbito de la economía colaborativa porque es posiblemente el motivo más destacado en la selección de alojamiento por parte de los huéspedes y también afecta significativamente a las ganancias de los anfitriones. Este hecho indica lo útil y valorado que puede ser analizar los factores que afectan al precio para poder guiar a los anfitriones y huéspedes hacia un aprovechamiento eficiente de la economía colaborativa.

El precio de Airbnb en Valencia es la variable elegida para el análisis estadístico. La unidad monetaria en la que está expresado en Inside Airbnb (página web de descarga de los datos) es en dólares (\$).

El precio medio por noche de los alojamientos Airbnb en Valencia es de 63\$ con una variabilidad que puede ir desde el mínimo de 9\$ al máximo de 818\$ (Tabla A2 - Anexo). Para comprobar la distribución de la variable dependiente, se han elaborado dos histogramas (Figura A2 - Anexo). El primer histograma, muestra una distribución

claramente sesgada a la izquierda, debido a la gran cantidad de precios concentrados en valores bajos y la existencia de precios altos de forma exclusiva para escasos alojamientos. Una alternativa que puede ayudar a normalizar esta distribución e incluir cierta no-linealidad es la conversión de la variable numérica a logaritmos. Tras transformarla (\log_price), se ha ejecutado un segundo histograma donde la distribución de la variable se ha normalizado y se puede apreciar su forma Gaussiana. El precio en unidad logarítmica es la variable que se toma para analizar las posibles características espaciales que contiene. En el modelo de regresión lineal se podrá ver con más detalle su mayor eficacia.

Al introducir los factores espaciales en el análisis, se ha de tener en cuenta la existencia de algún posible patrón de autocorrelación espacial. La autocorrelación espacial mide la relación entre una variable específica de una serie de objetos cercanos en comparación con otros objetos que también se encuentran cercanos entre ellos, basada en las ubicaciones y valores de las entidades de forma simultánea. Como dijo el geógrafo Waldo R. Tobler en la primera ley de geografía (1970): “Todas las cosas están relacionadas, pero las cosas cercanas están más relacionadas que las distantes”. Es un paso necesario que se debe realizar porque si hay evidencia de que existe autocorrelación espacial en los datos, las observaciones no están consideradas independientes unas de las otras y esto provocará una alteración en la eficacia del modelo OLS.

Un estadístico muy utilizado para la comprobación de autocorrelación espacial es Moran I. Este estadístico se encarga de evaluar si hay un patrón espacial de agrupación en clusters, disperso o es aleatorio. Analizando los resultados (Tabla A3 - Anexo), el valor-p indica que se rechaza la hipótesis nula (los precios Airbnb están distribuidos aleatoriamente en el área territorial de Valencia), por lo que la hipótesis alternativa confirma que existe autocorrelación espacial en la variable dependiente. El valor que expresa el estadístico de Moran I es positivo (0,036053), lo que indica que tiende a formar clusters en el espacio. Además, ejecutando el contraste local de asociación espacial (estadístico local de Moran I) y elaborando un mapa que aplica los resultados locales (Figura A3 - Anexo) se muestra la existencia de clusters, aunque con un problema: no diferencia si las agrupaciones son de valores altos o bajos.

Para solucionar este asunto, se ha creado un mapa de clusters LISA, un tipo de mapa en el que se pueden identificar las relaciones de valores altos/altos, altos/bajos, bajos/alto o bajos/bajos destacados que se comparten entre vecinos. Las regiones coloreadas son aquellas que contribuyen de forma significativa a la autocorrelación de la variable.

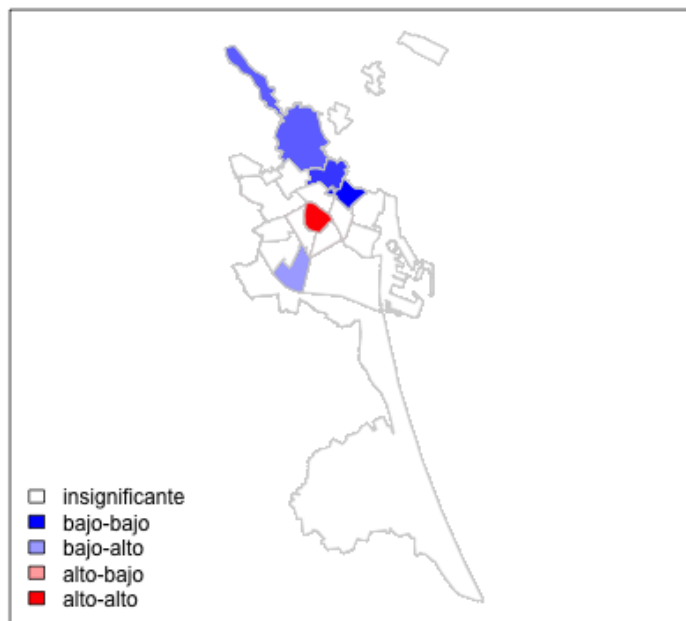


Figura 2 – Mapa de clusters LISA

En la Figura 2 se pueden diferenciar aquellos distritos en los que existen relaciones significativas entre los vecinos. Los distritos significativos son los que adoptan valores positivos en el mapa de Moran I representados por colores verdes (Figura A3 – Anexo). Ciutat Vella aparece claramente marcado en color rojo, representando una relación de valores altos-altos entre los alojamientos vecinos de la zona. En cambio, en los distritos Rascanya y Benimaclet se distinguen relaciones de valores bajos con valores bajos en ese territorio. Poblat del Nord y Jesús presentan una relación mixta de valores bajos con altos.

Aparentemente, existen patrones de precios estadísticamente significativos formando clusters en algunas zonas de la ciudad. El análisis hot spot es un método que sirve para observar los vecinos dentro de una proximidad definida e identificar dónde se forman clusters de valores altos o bajos de una variable en el espacio. Los precios altos rodeados por otros también altos están considerados estadísticamente significativos y son llamados hot spot. Por otro lado, los cold spots también son estadísticamente significativos con la diferencia de que representan una agrupación de observaciones con valores bajos del precio.

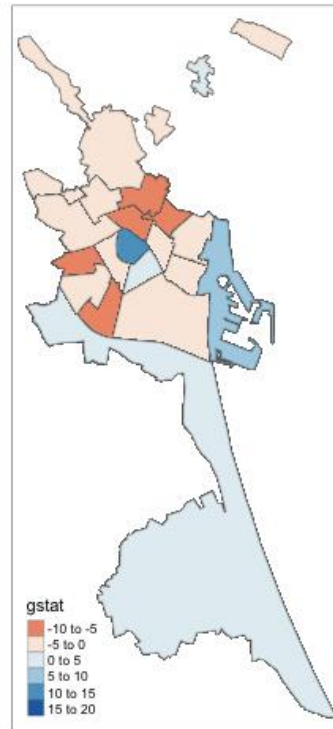


Figura 3 – Mapa Hot Spot de la variable “log_price”

En la Figura 3 se ha utilizado el estadístico Getis-Ord Gi (gstat) para poder hallar conjuntos de valores altos o bajos en el precio. RStudio ha permitido implementar una función que define una cantidad de vecinos basada en la proximidad. Se aprecia de forma clara que el distrito centro de Valencia (Ciutat Vella) es el hot spot más significativo debido a la proximidad a gran cantidad de lugares de interés. Sin embargo, Poblats Marítims al ser el distrito en el que se encuentra la playa principal de la ciudad (playa de la Malvarrosa) y la gran afluencia de turismo de sol y playa, también identifica un hot spot en esa zona. Por otra parte, los cold spots más destacados se encuentran en los barrios de La Saldia, Rascanya, Benimaclet, l'Olivereta y Patraix, donde se observan los niveles más bajos del precio de los Airbnb de la capital valenciana. Los demás distritos se agrupan en cold spots de valores más moderados.

Este fenómeno también se puede hallar entre los atributos de los alojamientos y en las variables creadas. Obviamente, las variables que describen la distancia al centro y a la playa, guardarán una alta autocorrelación espacial porque representan distancia fija de cada observación a un punto común.

Con este análisis, se han podido ver las diferencias locales y tipos de agrupación del precio que existen en los diferentes distritos, confirmando de esta forma la no-estacionariedad de las observaciones. Fotheringham, Brunson y Charlton (2002) aclaran

que la no-estacionariedad tiende a ser una característica de los procesos que suponen análisis sociales y declaran que la medida de una relación depende en parte de dónde se tome esa medida. En el caso de procesos espaciales se menciona como no-estacionariedad espacial. Estudios anteriores han examinado la no-estacionariedad espacial en los modelos de precios hedónicos de forma explícita, como es el caso de Pavlov (2000), aunque la imposibilidad de reunir todos los determinantes espaciales de los precios de las viviendas, por lo que casi siempre existe un sesgo de especificación errónea en el modelado estadístico de estos estudios.

5.2 Modelos de regresión lineal (OLS)

En este estudio se investiga acerca de hallar el mejor modelo de regresión lineal (múltiple al contener varios predictores) y utilizarlo como punto de partida del proceso. El objetivo de su ejecución es entender las relaciones entre los factores escogidos en el área de estudio. Para ello, se estudian las relaciones entre la variable dependiente (precio) y las variables independientes, que se irán incorporando a los modelos para evaluar su impacto.

Este método es fundamental, ya que trata de hallar las relaciones más adecuadas entre las variables que explicarán la variabilidad del precio de los alojamientos. A través del proceso se ejecutan varios modelos de regresión con el objetivo de comparar los resultados que refleja incluyendo nuevas variables o simplemente realizando transformaciones en sus unidades de medida.

Para validar su funcionamiento, se han creado un subconjunto de entrenamiento y un subconjunto de prueba. El primero de ellos representa el 80% de las observaciones, una muestra considerable para poder generar resultados estadísticamente significativos. Además, para garantizar que sea representativo de todo el conjunto de datos, se ha decidido utilizar los 85 barrios de la ciudad como grupos a los que aplicar este porcentaje y así cubrir subdivisiones más pequeñas que los distritos en el espacio. Los barrios son las áreas tomadas porque los datos no proporcionan otra división que separe el territorio en áreas más pequeñas. El segundo subconjunto representa el 20% restante y se utiliza para probar el modelo de entrenamiento y obtener los resultados del error cuadrático medio (RMSE) y el coeficiente de determinación (R-cuadrado). El RMSE determina la media de los errores (diferencia entre el estimador y lo que se estima) al cuadrado, por lo que cuanto menor sea su valor, mejor será para el modelo. Por otro lado, el R-cuadrado

mide el porcentaje de variación que suponen los predictores del modelo sobre la variable de estudio.

En el primer modelo, se utiliza el precio en sus unidades originales como variable dependiente junto a las siguientes variables independientes (Tabla A4 – Anexo): tipo de habitación (room_type), número de huéspedes (accommodates), distancia lineal al centro en kilómetros (distance_center_km) y distancia lineal a la playa más cercana en kilómetros (distance_beach_km). Se prevé que los resultados que indican el RMSE y el R-cuadrado pueden mejorar, por lo que se opta por realizar una transformación logarítmica del precio para incluir cierta no-linealidad en la variable dependiente del modelo y probar su eficacia.

El siguiente modelo obtiene unos resultados significativamente mejores. El hecho de haber transformado la variable dependiente a unidad logarítmica no solo ha supuesto una normalización de su distribución, sino que también ha mejorado el modelo. Al comprobar esta clara mejora, en el modelo 3 se han transformado las variables de distancia al centro y a la playa también a logaritmos (log_distance_center y log_distance_beach). El resultado se ha visto mejorado en aproximadamente un 2% con respecto al anterior modelo.

Tras este análisis, se prueba incluir las variables de los niveles de demanda de los equipamientos municipales. En la matriz de correlación de Pearson (Figura A1 – Anexo) se puede comprobar que existe correlación entre la variable dependiente y las variables de niveles de demanda, pero a la vez se puede ver la alta correlación negativa que éstas suponen con relación a la variable de distancia al centro de la ciudad. Este hecho puede conllevar a problemas de multicolinealidad en la regresión.

El Factor de Inflación de la Varianza (VIF) refleja los problemas de colinealidad que puede tener el modelo. Un valor 1 supone una ausencia total de colinealidad, valores cercanos a 5 pueden significar un efecto de colinealidad en la regresión y valores mayores que 5 causan preocupación. Los resultados del VIF en el modelo que incluye la variable de demanda muy alta (VHigh), muestran valores cercanos a 5 en las variables “log_distance_center” y “VHigh”. De esta forma, parece confirmarse la multicolinealidad entre las dos variables de las que se sospechaba. Esto puede deberse a que los monumentos y sitios de interés seleccionados se encuentran en mayor medida alrededor del centro de la ciudad y, por lo tanto, a mayor distancia al centro menor cantidad de monumentos (alta correlación negativa).

Teniendo en cuenta los resultados que ofrece el modelo que incluye la variable de demanda, no supone una mejora relevante en los resultados, por lo que se decide eliminar las variables de demanda del análisis de regresión. Además, los problemas de colinealidad que presentan no suponen un incentivo para su inclusión como predictor. Considerando los resultados obtenidos en los 4 modelos analizados, el modelo que se escoge es el modelo 3.

La existencia de observaciones con valores atípicos es otro asunto que se considera para validar la efectividad del método OLS. Existen dos enfoques en la consideración de estos valores. En el primero de ellos, se eliminan aquellas entidades que contienen valores fuera de lo normal si se considera todo el conjunto de datos. El segundo enfoque los incluye en el modelo para tratar de explicarlos. En este estudio se ha optado por el primer enfoque, ya que está encaminado a la predicción de los precios de alojamientos de categoría estándar y se ha prescindido de aquellos catalogados como “de lujo” al ser una minoría que puede sesgar el modelo. Para ello, se han elaborado dos gráficos (Figura A4 - Anexo) donde se observa el primero con los datos atípicos (puntos rojos) y el segundo muestra la limpieza de éstos.

Los nuevos datos incluyen, por ende, nuevas características (Tabla A5 - Anexo). Con estos datos, se ha elaborado de nuevo un modelo de regresión lineal con las variables del modelo 3. Los resultados indican una mejora muy notable del R-cuadrado al sufrir un incremento de un 9% (de 53% a 62%) con respecto al modelo que incluía todas las observaciones. Además, el RMSE ha disminuido de 0,4593 a 0,3957.

El modelo final (modelo 5) se describe en la Tabla 3, donde aparecen los coeficientes para cada variable independiente incluida.

Modelo 5: $\log_price = \alpha + \beta_1 \text{room_type}(\text{shared}) + \beta_2 \text{room_type}(\text{private}) + \beta_3 \text{accommodates} + \beta_4 \log_distance_center + \beta_5 \log_distance_beach$

Variable explicativa	Coeficiente	Error estándar	t-valor	p-valor
constante	5,1912	0,1134	45,7720	0,0000
room_type (private)	-0,6558	0,0157	-41,7280	0,0000
room_type (shared)	-0,8331	0,0939	-8,8720	0,0000
accommodates	0,1144	0,0034	33,5780	0,0000
log_distance_center	-0,1287	0,0081	-15,8170	0,0000
log_distance_beach	-0,0768	0,0083	-9,2920	0,0000

Tabla 3 - Resumen de los resultados estimados del modelo OLS final.

Se puede observar (Tabla 3) que la variable “room_type” está dividida en “private” y “shared”. Esto se debe a la presencia de una variable cualitativa separada en una serie de factores con el objetivo de diferenciarlos y obtener los resultados para cada uno de ellos. Este tipo de variables se llaman “dummy” y se encargan de dar valores de 0 y 1 a los factores de una variable, tratándolas como nuevas variables dentro del modelo. En este caso, los coeficientes de las dos que aparecen se relacionan con la variable omitida “room_type (entire home/apt)”, haciendo referencia a los tipos de alojamiento que la totalidad del espacio de la casa ofertada.

Los resultados indican que todos los coeficientes son estadísticamente significativos ($p\text{-valor} < 0,05$). Estos coeficientes tienen distintos impactos sobre el precio. Las habitaciones privadas y compartidas suponen una disminución del precio (con respecto a la oferta de la casa entera) y la distancia al centro y a la playa también disminuyen el precio conforme aumentan sus valores, ya que son dos ubicaciones muy demandadas por los turistas a la hora de tener mayor accesibilidad a lugares de interés en sus periodos vacacionales. En cambio, la variable de la capacidad de huéspedes (accommodates) tiene un impacto positivo sobre el precio, donde una mayor capacidad de personas en el alojamiento supone un incremento del precio final. Todos los signos de los coeficientes estimados tienen sentido a la hora de aplicar el nivel de precio y se podía prever, pero con el análisis ejecutado se puede comprobar en qué medida suponen un impacto sobre la variable de estudio.

Mediante los gráficos del modelo con todas las observaciones y el modelo sin los valores atípicos (Figuras A5 y A6 – Anexo) se puede comparar su bondad de ajuste. La diferencia más apreciable entre ambos modelos se observa en el gráfico “Normal Q.Q”, el cual demuestra el grado en que los residuos se distribuyen de forma normal. Los residuos normales se ajustan a la línea diagonal intermitente trazada y se aprecia cómo el modelo sin valores atípicos ofrece un mejor ajuste de las observaciones a la línea de referencia. Para confirmar estos resultados, se ha elaborado un histograma que muestra la normalidad que tienen los residuos del modelo final con relación al modelo que incluye todas las entidades (Figura A7). A pesar de que no es un requisito fundamental en la corrección del modelo, en los resultados de ambos (con datos atípicos y sin atípicos) se ha podido observar cómo el R-cuadrado y RMSE han mejorado significativamente.

5.3 Modelo de regresión ponderada geográficamente (GWR)

Previamente, se ha comprobado que las relaciones entre las observaciones no son estacionarias, significando esto que los modelos OLS pueden causar problemas en la interpretación de las estimaciones paramétricas, ya que asumen las relaciones entre las entidades como estacionarias (Brunsdon, Fotheringham y Charlton, 1998). Es por eso, que la regresión ponderada geográficamente reconoce la existencia de variaciones a través del espacio dependiendo de la localización de cada entidad, solucionando las cuestiones espaciales que el modelo tradicional no es capaz de explicar.

En anteriores estudios similares como en el de Zhang et al. (2017), se ha comprobado que el GWR tiene mejor bondad de ajuste que el OLS. A pesar de ello, se ha empezado prediciendo el mejor modelo OLS para poder aplicar con posterioridad el método GWR al mismo conjunto de variables explicativas potenciales y solucionar los problemas de variación espacial.

Para elaborar el modelo, se ha utilizado la librería “spgwr” de R. Cabe destacar el elevado tiempo de ejecución de la función de GWR (alrededor de una hora) y la complejidad computacional en el desarrollo del proceso. Para validar el correcto funcionamiento del método, se han utilizado también una muestra de entrenamiento y una muestra de prueba de los datos, como en OLS. Previamente a la ejecución del modelo GWR, es necesario calcular el ancho de banda (bandwidth) del kernel para cada observación.

Anteriormente, se ha aclarado que en este estudio se va a utilizar el cálculo de un kernel gaussiano adaptativo por su mejor aplicación a un análisis de estas características, buscando la cantidad óptima de vecinos para cada entidad en el espacio. El criterio de evaluación ha sido la validación cruzada (cross-validation), con el que se ha obtenido el porcentaje global de observaciones que proyecta la puntuación de validación cruzada más baja. En este caso, la aplicación del ancho de banda adaptativo indica que el número óptimo de entidades vecinas es 40.

Tras la obtención del ancho de banda, se ha aplicado a la ejecución del modelo de regresión espacial ponderada geográficamente sobre la muestra de entrenamiento y se han obtenido una serie de resultados con los que se pueden identificar grandes variaciones en los coeficientes por cuartiles debido a las variaciones espaciales (Tabla 3).

Variable explicativa	Mínimo	Cuartil 1	Mediana	Cuartil 3	Máximo	Global
constante	-39,9552	-1,7571	6,0422	12,7785	159,6560	5,1912
room_type (private)	-1,1548	-0,7335	-0,6396	-0,5360	-0,1227	-0,6558
room_type (shared)	-4,7954	-0,8911	-0,6830	-0,3999	0,5607	-0,8331
accommodates	0,0070	0,0919	0,1082	0,1242	0,1850	0,1144
log_distance_center	-3,1140	-0,2658	-0,0632	0,2783	6,7475	-0,1287
log_distance_beach	-23,5896	-0,7694	-0,1237	0,2549	4,0142	-0,0768

Tabla 3 - Resumen de los resultados estimados del modelo GWR

La Tabla 3 muestra los coeficientes que adoptan las variables independientes diferenciadas en cuartiles, donde el mínimo representa la observación con el valor más bajo, el primer cuartil es el valor que deja el 25% de las observaciones por debajo de él, la mediana señala el valor que divide los datos en dos partes iguales, el tercer cuartil describe que el 75% de los datos es menor o igual a este valor y el máximo es el mayor valor que adopta ese coeficiente. La última columna muestra los coeficientes globales del modelo de regresión de partida.

Estas variaciones que se observan en la tabla confirman la heterogeneidad espacial que las variables explicativas experimentan a través del espacio, justificando de esta forma la mayor eficacia de la regresión ponderada geográficamente con respecto a la regresión lineal tradicional en el análisis del precio de los alojamientos de la ciudad. Los coeficientes globales de las variables se corresponden con el mismo signo que en el OLS, aunque “room_type (shared)”, “log_distance_center” y “log_distance_beach” llegan a presentar tanto coeficientes positivos como negativos según la entidad del espacio que se estudie.

La aplicación de la muestra de prueba al modelo ha confirmado la mejora de resultados que proporciona el GWR frente al OLS. Entre los resultados a destacar en la ejecución del modelo, el RMSE ha disminuido significativamente desde 0,3957 hasta 0,0037. El R-cuadrado también indica que se ha mejorado la explicación de la varianza del precio por parte de los predictores, ascendiendo hasta aproximadamente un 65% (0,6533) frente al 62% (0,6205) del método OLS. Sin embargo, este resultado es global y tiene variaciones locales dependiendo de la zona. En la Figura 4 se observa la distribución del R-cuadrado local para cada distrito, con valores comprendidos entre 0 y 1 para comprobar la bondad de ajuste del modelo de regresión local de los valores observados.

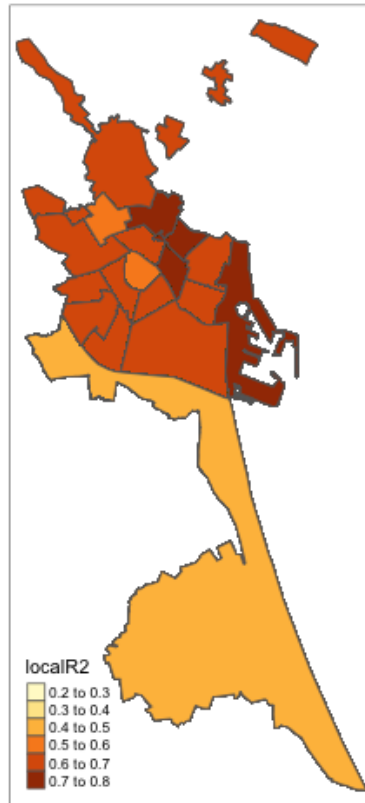


Figura 4 – Distribución del R-cuadrado por distritos

El mapa del R-cuadrado local mide la proporción de la varianza en los datos observados explicados por el modelo. Las variaciones en este estadístico pueden ser computadas para dar un sentido a cómo de bien un modelo local puede replicar los datos guardados en la vecindad del punto de regresión.

Fotheringham, Brundson y Charlton (2002) definen la fórmula del R-cuadrado local de la siguiente forma:

$$r_i^2 = (TSS^w - RSS^w) / TSS^w$$

donde TSS es la suma total de cuadrados ponderada geográficamente, definida como

$$TSS = \sum_j w_{ij} (y_j - \bar{y})^2$$

y RSS es la suma de cuadrados residual ponderada geográficamente, definida como

$$RSS = \sum_j w_{ij} (y_j - \hat{y}_i)^2$$

donde w_{ij} describe el peso del punto j en el punto de regresión i .

Como se puede observar, quitando Poblats del Sud, todos los distritos explican entre el 50% y el 80% la varianza del precio. El hecho de que las zonas superen mayoritariamente el 50% convierte el modelo en un método considerablemente eficaz a la hora de estimar los precios Airbnb en Valencia.

En cuanto a lugares a destacar, se podrían nombrar los distritos de Rascanya, Benimaclet, el Pla del Real y Poblats Marítims, con valores que oscilan entre el 70% y 80%. Los distritos que se encuentran en las zonas del Oeste, Norte y alrededor del centro de Valencia presentan un R-cuadrado superior al 60%, mientras que Ciutat Vella (centro de la ciudad) y Benicalap experimentan los valores más bajos de la ciudad (50%-60%), quitando Poblats del Sud que se encuentra entre 40% y 50%.

Por otra parte, es importante analizar el impacto local que tienen las variables en las distintas zonas, ya que en la tabla también se observa cómo presentan variaciones en el espacio que incluso pueden llegar a cambiar el signo del coeficiente de algunos predictores. Para ello, se han creado 5 gráficos correspondientes a los coeficientes que las variables explicativas adoptan en los resultados de GWR con diferencias notables según el distrito correspondiente.

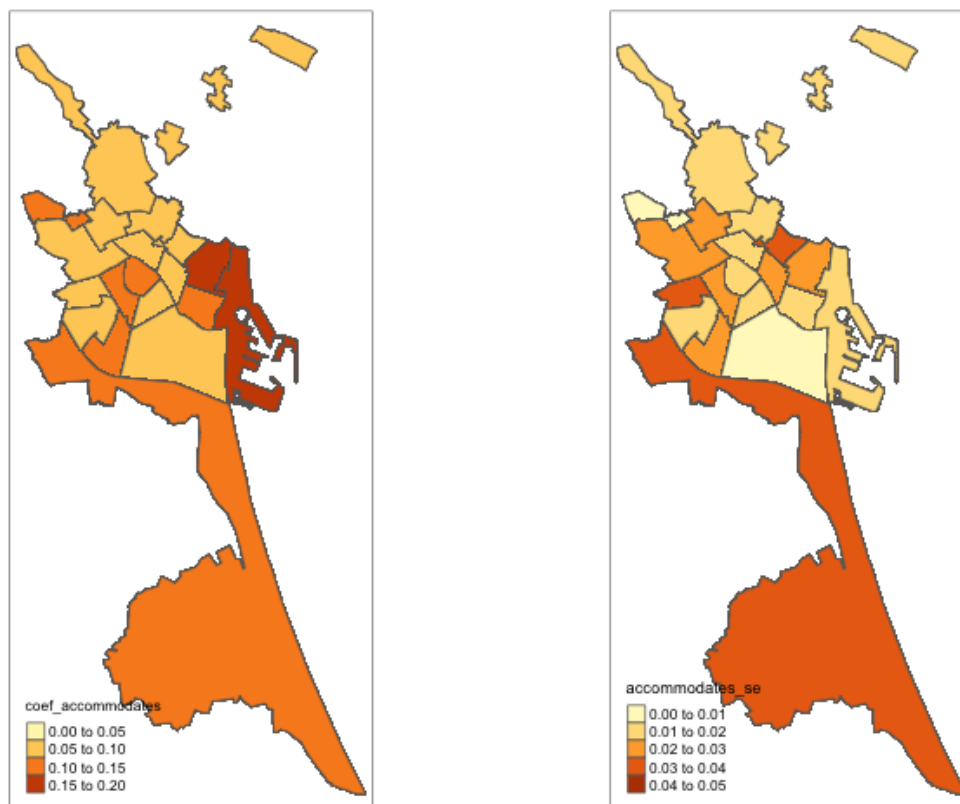


Figura 5 – Mapas de coeficientes y errores estándar de “accommodates” por distritos

Según los resultados de la Figura 5, la capacidad de huéspedes que el alojamiento oferta presenta coeficientes positivos en todo el territorio, justificando de esta forma la previsión relativamente obvia de que cuanto mayor sea la capacidad de huéspedes, mayor será el precio.

Entre los distritos a destacar, los que experimentan un mayor impacto de esta variable son los que contienen alojamientos cercanos a la playa de la Malvarrosa. El centro de Valencia también tiene un impacto de la capacidad de huéspedes mayor que los demás distritos que le rodean, aunque con valores algo más bajos que los distritos costeros. El mapa de errores estándar (se) muestra unos valores heterogéneos pero muy cercanos a 0 en todo el territorio y por debajo de los valores que adoptan los coeficientes, pareciendo ser que se aprueba la confianza de estos resultados obtenidos.

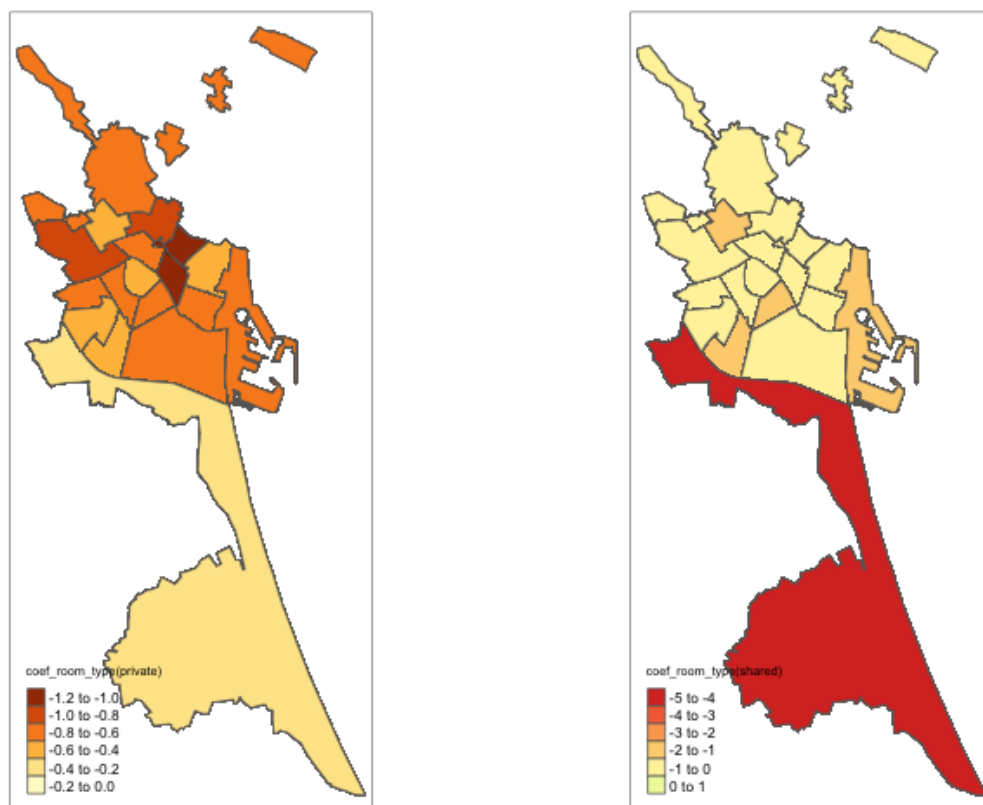


Figura 6 – Mapas de coeficientes de “room_type” por distritos

En la Figura 6 aparecen dos mapas que representan los coeficientes de “room_type” separados en los tipos de habitaciones privadas y compartidas. El primer mapa hace referencia a los coeficientes que supone la oferta de una habitación privada en un alojamiento con más huéspedes. Como el tipo de habitación que el modelo toma como referencia es “apartamento entero”, el impacto que tiene sobre el precio la oferta de una

habitación privada será negativo en todo el territorio valenciano al ofrecer un espacio e intimidad claramente menor. Los distritos en los que este tipo de oferta supone una disminución más significativa del precio son el Pla del Real y Benimaclet. Rascanya y Campanar presentan un impacto algo menor pero igualmente considerable con relación a los demás distritos que presentan coeficientes algo más bajos. Poblats del Sud no presenta un coeficiente elevado debido a su disminuida oferta de este tipo de habitación en el distrito, con unos precios moderados sin demasiada variabilidad.

El efecto de la oferta de habitaciones compartidas sobre el precio de los alojamientos es muy similar a la de las habitaciones privadas. Todos los distritos se encuentran entre valores 0 y -2 a excepción de Poblats del Sud, que experimenta un coeficiente altamente negativo debido a que en este distrito tan sólo se oferta una habitación compartida y cae todo el peso de la zona sobre ese valor que, en este caso, es muy bajo.

Parece ser que se puede confirmar la confianza de estos resultados teniendo en cuenta los mapas de errores estándar de ambas variables en el anexo (Figura Ax) que muestran unos bajos valores de error, a excepción de Poblats del Sud debido a su escasez de oferta.

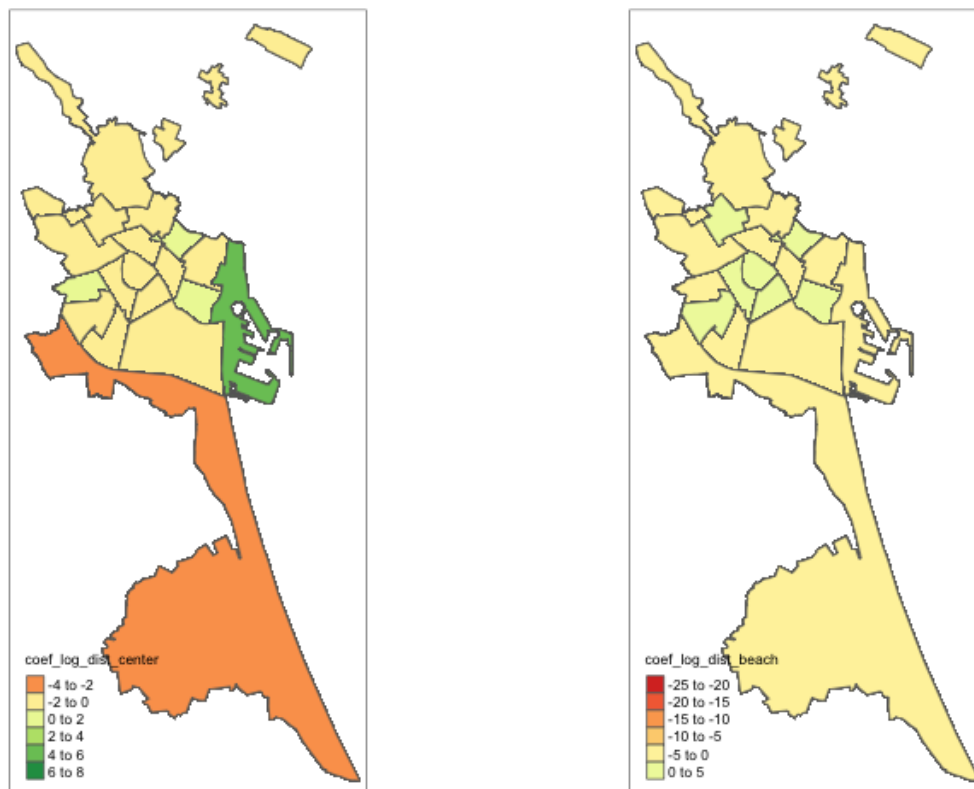


Figura 7 – Mapa de coeficientes de “log_distance_center” y “log_distance_beach” por distritos

La Figura 7 distingue los coeficientes locales que adoptan las variables de distancia creadas en el desarrollo del trabajo. En ambos mapas se puede apreciar que hay zonas que presentan coeficientes positivos y otras negativos, a pesar de que los coeficientes globales de las dos variables son negativos. Esto se debe a que las dos variables representan las dos zonas con los precios más elevados (media de 73\$ y 68\$ en el centro y en la costa de la Malvarrosa, respectivamente, frente a la media general de 60\$) debido a las ventajas culturales y ociosas que presentan. El distrito con el coeficiente negativo más impactante de la distancia al centro sobre los precios de sus Airbnb es Poblats del Sud, ya que se encuentran a una distancia muy considerable de las zonas más demandadas de la ciudad. Poblats Marítims presenta un coeficiente positivo alto debido a que su distancia lejana al centro se contrarresta con su distancia muy cercana a las playas. Los demás distritos presentan unos coeficientes negativos bastante moderados, entre 0 y -2.

En el caso de las distancias a la playa más cercana, la ciudad muestra unos coeficientes negativos entre 0 y -5 en la mayoría del territorio. Cabe mencionar a Benimaclet y Camins al Grau como únicos distritos que presentan coeficientes positivos tanto en la variable de distancia al centro como a la playa. Esta característica inusual que presentan estas zonas probablemente se debe a que son distritos que se encuentran en un punto medio entre el centro de la ciudad y la playa de la Malvarrosa, siendo así dos de los distritos con mejor ubicación tanto para ir al centro como a la costa.

De igual forma que las variables de “room_type”, se pueden visualizar los mapas de errores estándar (Figura A9 - Anexo) de las variables de distancia para comprobar sus bajos valores y, por lo tanto, parece confirmar la confianza que se preveía sobre los resultados de sus coeficientes.

6 CONCLUSIONES

El empleo del software RStudio ha permitido afianzar los conocimientos adquiridos del mismo durante el grado en Administración y Dirección de Empresas para un mayor entendimiento del análisis y muestreo de datos. Además, se ha profundizado en el aprendizaje del propio software y de sus herramientas integradas al utilizarlo de forma exhaustiva durante todo el desarrollo del proceso estadístico.

El análisis de los datos de Airbnb en Valencia ha dado lugar a extraer varias conclusiones. Ciutat Vella y Poblats Marítims tienen un claro protagonismo a la hora de

calcular sus precios medios, ya que son los distritos en los que existe una mayor oferta debido a sus características generales de interés turístico.

Con el objetivo de hallar indicadores que explican la variable del precio, se han incluido predictores cuantitativos y cualitativos en una serie de modelos de regresión lineal a través del uso de muestras de entrenamiento y prueba para la validación de su función. Además, la creación de nuevas variables que contienen características espaciales ha ayudado a analizar su impacto en el modelo. Considerando los resultados del RMSE y R-cuadrado obtenidos en los modelos OLS, se ha comprobado cómo la inclusión de algunas de las variables creadas provocaba problemas de multicolinealidad que conllevaban a la falta de eficacia del modelo final. A pesar de esto, se han podido añadir finalmente las variables que calculaban la distancia lineal al centro y a la playa más cercana, transformadas a unidad logarítmica por la mejora de resultados que proporcionaba.

Para comprobar la estacionariedad de las observaciones se ha elaborado un análisis de la autocorrelación espacial del precio. El uso del estadístico de Moran I y su interpretación local gráfica ha dado lugar a la posibilidad de realizar un mapa de clusters LISA para visualizar el tipo de relación entre los vecinos de un mismo distrito, además de un mapa hot spot de la variable dependiente donde se ha podido comprobar visualmente que las zonas de precios más elevados son el centro de la ciudad y la costa de la Malvarrosa.

La existencia de autocorrelación espacial ha demostrado que los alojamientos guardan relaciones de precio entre ellos, principio que altera la credibilidad del modelo tradicional. Como esperaba, los resultados obtenidos tras la ejecución del modelo GWR con ancho de banda adaptativo han ofrecido una mejor explicación de la varianza del precio que el modelo de regresión lineal final (65% frente al 62% en R-cuadrado).

Los resultados locales que se han obtenido a partir del modelo GWR han permitido elaborar un mapa del R-cuadrado local para comprobar la varianza del precio en los distintos distritos que tienen diferentes niveles de impacto según la zona en la que se encuentren los alojamientos. Rascanya, Benimaclet, el Pla del Real y Poblets Marítims son los distritos que presentan los resultados más elevados del R-cuadrado local, con unos valores comprendidos entre 0,7 y 0,8 sobre 1. El resto del territorio presenta unos valores algo más bajos, aunque todos por encima de 0,5 exceptuando Poblets del Sud por su escasez de oferta de alojamientos.

En cuanto a los coeficientes locales, se ha visto que la variable que describe la capacidad máxima de huéspedes en el alojamiento tiene un impacto muy similar en toda la ciudad, mientras que las variables de distancia incluyen un rango de valores más

extenso. Esto es debido a que, generalmente, las distancias que multiplican a los coeficientes pueden ser valores relativamente pequeños si se encuentran muy cerca de la zona, pero hay alojamientos que se ubican a una larga distancia tanto del centro como de las playas, como ocurre en el caso de Poblat del Nord.

Con el análisis finalizado, los resultados obtenidos por el modelo final han servido de guía orientativa en el descubrimiento de variables que tienen un impacto directo sobre los precios, según las descripciones de los métodos de regresión empleados. Se ha podido ver cómo la localización de las entidades adopta unos coeficientes distintos dependiendo de la zona en la que se encuentran, pudiendo haber representado de forma visual una aproximación en forma de intervalo de valores que describen los resultados medios por distritos.

7 BIBLIOGRAFÍA

Acosta Perera, A. (2017) *Análisis especial del Mercado de alojamiento de Airbnb en la isla de Gran Canaria en 2017*. Universidad de Las Palmas de Gran Canaria

Adamiak, C., Szyda, B., Dubownik, A. y García-Álvarez, D. (2019) *Airbnb Offer in Spain – Spatial Analysis of the Pattern and Determinants of Its Distribution*.

Barron, K., Kung, E. y Proserpio, D. (2017) *The Effect of Home-Sharing on House Prices and Rents: Evidence from Airbnb*.

De la Torre, N. y Toledo, C. (2019) “*Valencia pide ayuda a Europa para controlar la expansión de Airbnb*”, El Mundo, 25 de junio. Disponible en:

<https://www.elmundo.es/comunidad-valenciana/2019/06/25/5d10f38321efa086328b464e.html>

Edelman, B.G. y Geradin, D. (2016) *Efficiencies and regulatory shortcuts: how should we regulate companies like Airbnb and Uber?* Disponible en: <https://www-cdn.law.stanford.edu/wp-content/uploads/2017/11/19-2-4-edelman-geradin-final.pdf>

Fernandes, I. (2018) *Modelling the Airbnb listings' Price in Lisbon using spatial regressions*. Universidad Nova de Lisboa

Fotheringham, A.S., Brunsdon, C. y Charlton, M. (2002) *Geographically Weighted Regression: The analysis of spatially varying relationships*.

Fundació Visit València (2019) *Estadísticas de Turismo 2018*. Disponible en: <https://www.visitvalencia.com/sites/default/files/pdfs/fundacion/estadisticas/folleto-estadisticas-2018.pdf>

Gutiérrez, J., García-Palomares, J.C. y Daniel-Cardozo, O. (2012) *Regresión Geográficamente Ponderada (GWR) y estimación de la demanda de las estaciones del Metro de Madrid*. Disponible en: http://tig.age-geografia.es//2012_Madrid/ponencia2/GutierrezPuebla_final_par.pdf

Gutiérrez, J., García-Palomares, J.C., Romanillos, G. y Salas-Olmedo, M.H. (2017) *The eruption of Airbnb in tourist cities: Comparing spatial patterns of hotels and peer-to-peer accommodation in Barcelona*.

Guttentag, D. (2013) *Airbnb: Disruptive innovation and the rise of an informal tourism accommodation sector*.

https://www.researchgate.net/publication/271624904_Airbnb_Disruptive_innovation_and_the_rise_of_an_informal_tourism_accommodation_sector

Hosbec (2018) “*Airbnb dispara su negocio en Valencia hasta 45 millones*”, Expansión, 4 de abril. Disponible en:

<https://www.expansion.com/valencia/2018/04/04/5ac4fc6ee2704ef1118b4619.html>

Instituto Nacional de Estadística (2020) *Estadística de Movimientos Turísticos en Fronteras (FRONTUR)*. Disponible en:

<https://www.ine.es/daco/daco42/frontur/frontur1219.pdf>

Jiménez Horwitz, M. (2018) *La situación jurídica de la plataforma Airbnb en el marco de la economía colaborativa*. Universidad de Málaga. Disponible en:

https://www.uma.es/departamento-de-derecho-civil-derecho-ecclesiastico-del-estado-y-derecho-romano/navegador_de_ficheros/Proyecto-FEDER-Alojamientos-Turisticos/descargar/Articulos/Jimenez-Horwitz-M-AIRBNB.pdf

Koster, H.R.A., van Ommeren, J. y Volkhausen, N. (2018) *Short-term rentals and the housing market: Quasi-experimental evidence from Airbnb in Los Angeles*. Disponible en: https://cepr.org/active/publications/discussion_papers/dp.php?dpno=13094

Lansley, G. y Cheshire, J. (2016) *An Introduction to Spatial Data Analysis and Visualisation in R*. Disponible en:

<https://www.spatialanalysisonline.com/An%20Introduction%20to%20Spatial%20Data%20Analysis%20in%20R.pdf>

Lee, D. (2016) *How Airbnb Short-Term Rentals Exacerbate Los Angeles's Affordable Housing Crisis: Analysis and Policy Recommendations*. Disponible en: <http://blogs.ubc.ca/canadianliteratureparkinson/files/2016/06/How-Airbnb-Short-term-rentals-disrupted.pdf>

Pavlov, A.D. (2000) *Space-Varying Regression Coefficients: A Semi-parametric Approach Applied to Real Estate Markets*. Disponible en: <https://doi.org/10.1111/1540-6229.00801>

Quattrone, G., Proserpio, D., Quercia, D., Capra, L. y Musolesi, M. (2016) *Who Benefits from the "Sharing" Economy of Airbnb?* Disponible en: <https://dl.acm.org/doi/pdf/10.1145/2872427.2874815>

Statista (2019) *Las ciudades con más alojamientos de Airbnb*. Disponible en: <https://es.statista.com/grafico/16721/anuncios-airbnb-espana/>

Wang, S., Fang, C., Ma, H., Wang, Y. y Qin, J. (2014) *Spatial differences and multi-mechanism of carbon footprint based on GWR model in provincial China*. Disponible en: <https://link.springer.com/article/10.1007/s11442-014-1109-z>

Zhang, Z., Chen, R.J.C., Han, Lee D. y Yang, L. (2017) *Key Factors Affecting the Price of Airbnb Listings: A Geographically Weighted Approach*.

8 ANEXOS

Demanda Muy Alta (+2000 opiniones)	Demanda Alta (1000-2000 opiniones)
Mercado Central Parroquia de San Nicolás de Bari y San Pedro Mártir Lonja de la Seda Plaza del Ayuntamiento El Miguelete Ciudad de las Artes y las Ciencias El Oceanográfico Plaza del Mercado	Basílica de la Virgen de los Desamparados Estació del Nord Edificio de Correos y Telégrafos Plaza de la Reina Torres de Serranos L'Hemisfèric Bioparc L'Albufera Museo de Bellas Artes
Demanda Media (500-1000 opiniones)	Demanda Media-Baja (100-500 opiniones)
Iglesia del Patriarca o del Corpus Christi Estadio de Mestalla Torres de Quart Plaza de Toros de Valencia Plaza Redonda Museo Fallero de Valencia	Trinquet de Pelayo Iglesia de San Juan del Hospital Iglesia de Santa Catalina Palau de la Generalitat Museo y Colegio del Arte Mayor de la Seda

Tabla A1 – Clasificación de los equipamientos municipales según demanda

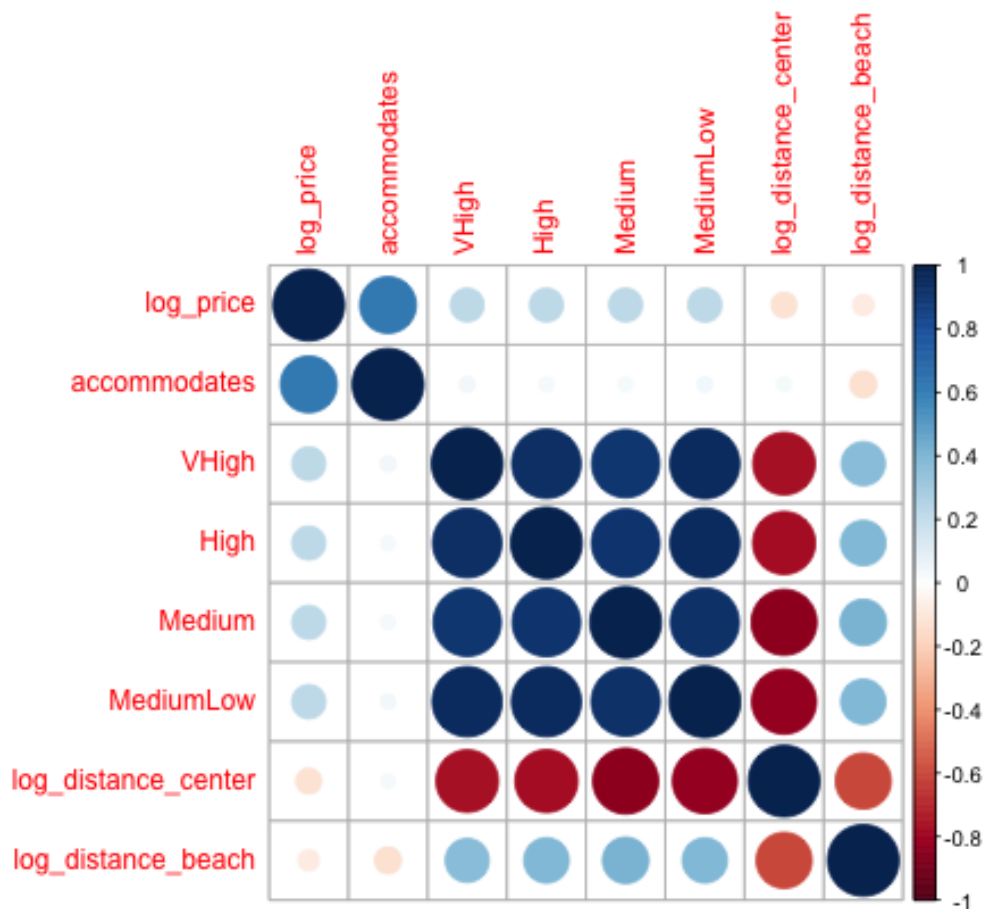


Figura A1 – Matriz de correlación Pearson

Variable	Mínimo	Cuartil 1	Mediana	Media	Cuartil 3	Máximo
final_price	9	35	54	63,4	75	818
accommodates	1	1	4	3,9	5	25
Vhigh	0	0	0	1,9	5	6
High	0	0	1	1,7	4	6
Medium	0	0	1	1,3	3	4
MediumLow	0	0	0	1,5	3	5
distance_center	21,4	838,4	1576	2141,9	2866,6	22765,2
distance_beach	0	2246	4294	3684	4820	10667

Tabla A2 – Resumen estadístico de las variables

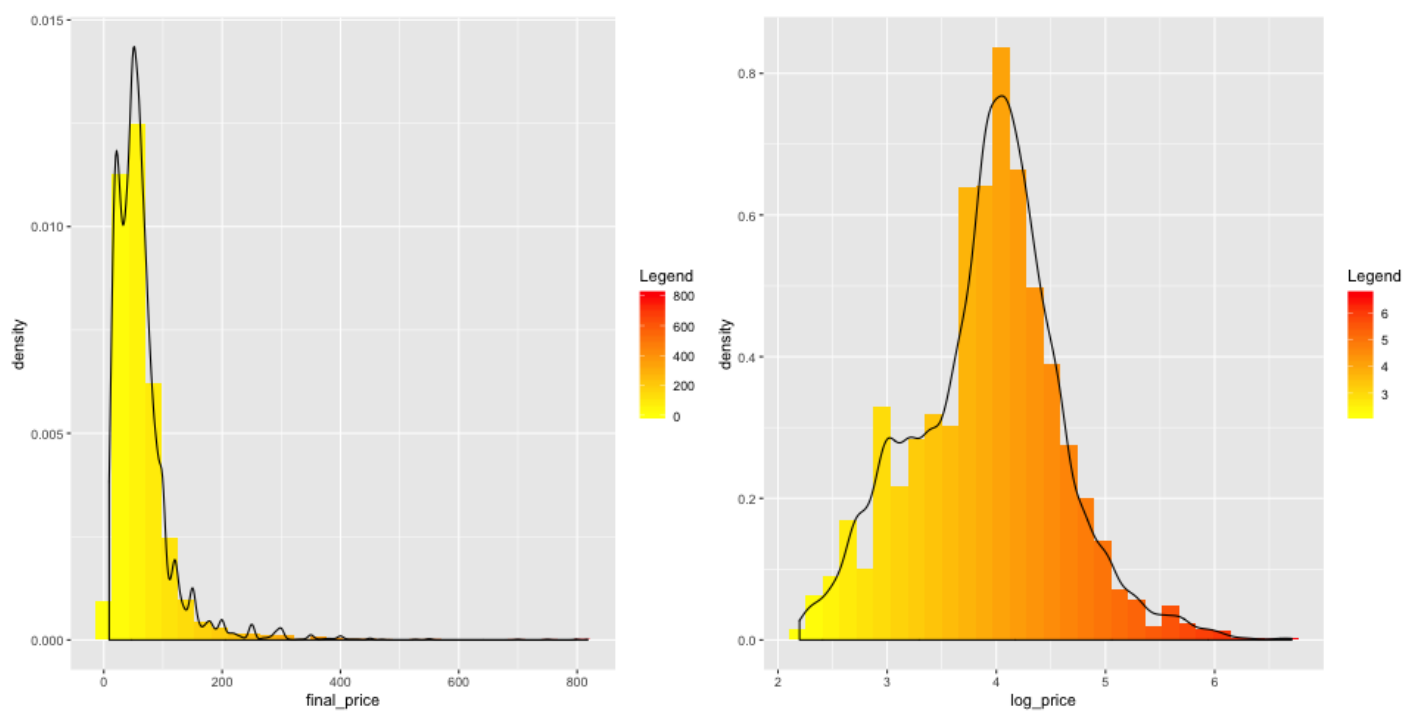


Figura A2 – Histogramas del precio en \$ y en unidad logarítmica

Estadístico de Moran I	0,036053
Esperanza	-0,000016
Varianza	0,000000
Desviación estándar	70,585
Valor-p	0,000000
Hipótesis alternativa	Mejor

Tabla A3 – Resultados globales de Moran I

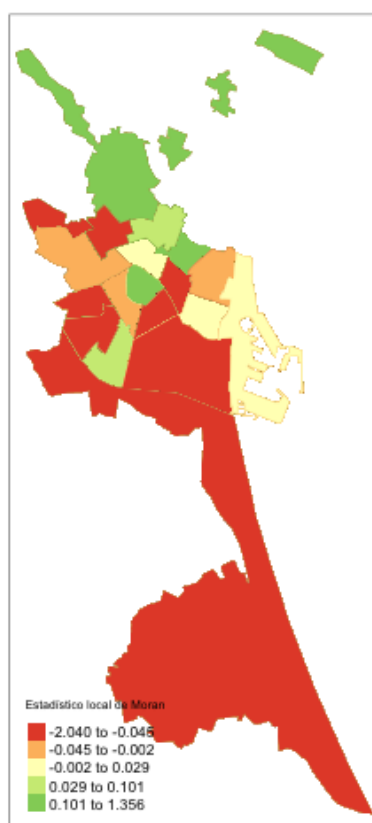


Figura A3 – Mapa del Estadístico Local de Moran I

Modelo	Variable dependiente	Predictores	RMSE	R-cuadrado
1	final_price	room_type + accommodates + distance_center + distance_beach	43,79	0,303
2	log_price	room_type + accommodates + distance_center_km + distance_beach_km	0,4671	0,5128
3	log_price	room_type + accommodates + log_distance_center + log_distance_beach	0,4593	0,5283
4	log_price	room_type + accommodates + log_distance_center + log_distance_beach + VHigh	0,4579	0,5322
5 (datos sin atípicos)	log_price	room_type + accommodates + log_distance_center + log_distance_beach	0,3957	0,6205

Tabla A4 – Modelos OLS

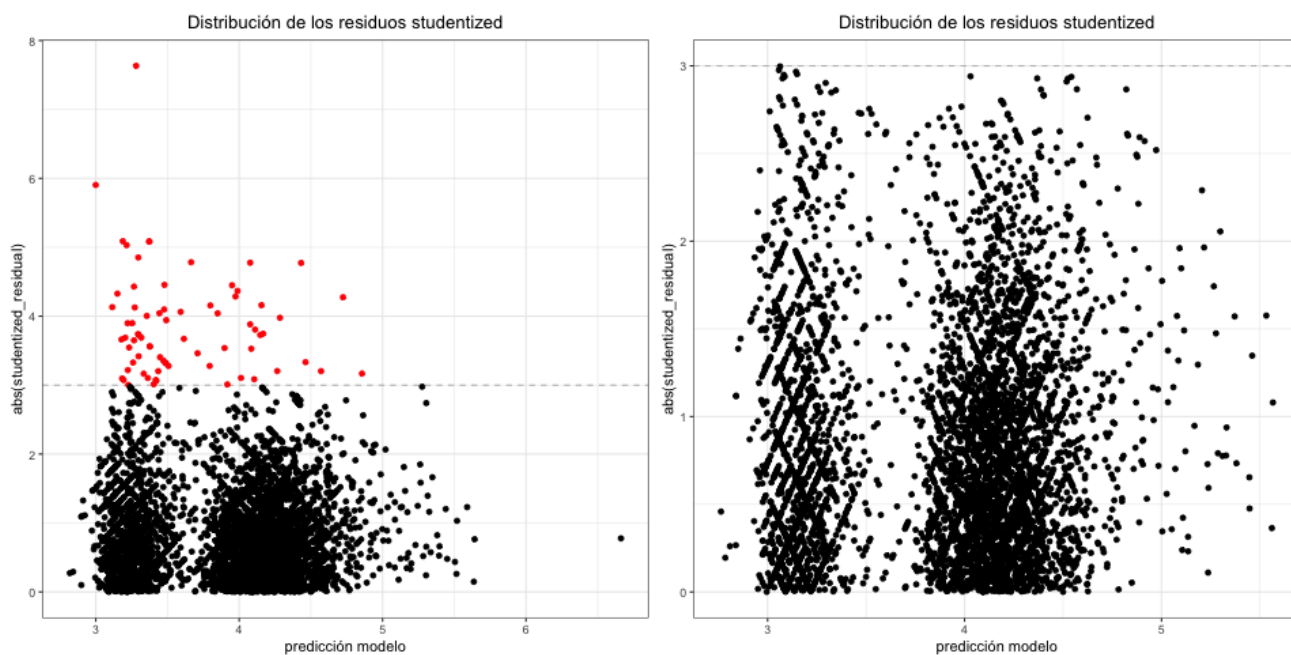


Figura A4 – Gráficos de los datos con valores atípicos y sin valores atípicos

Variable	Mínimo	Cuartil 1	Mediana	Media	Cuartil 3	Máximo
final_price	9	33	52	59,8	74	470
accommodates	1	2	4	3,9	5	16
Vhigh	0	0	0	1,9	5	6
High	0	0	0	1,9	5	6
Medium	0	0	1	1,3	3	4
MediumLow	0	0	0	1,5	3	5
distance_center	21,4	844,7	1586	2148,1	2873,9	22765,2
distance_beach	0	2231	4286	3676	4816	10667

Tabla A5 – Resumen estadístico de las variables sin valores atípicos

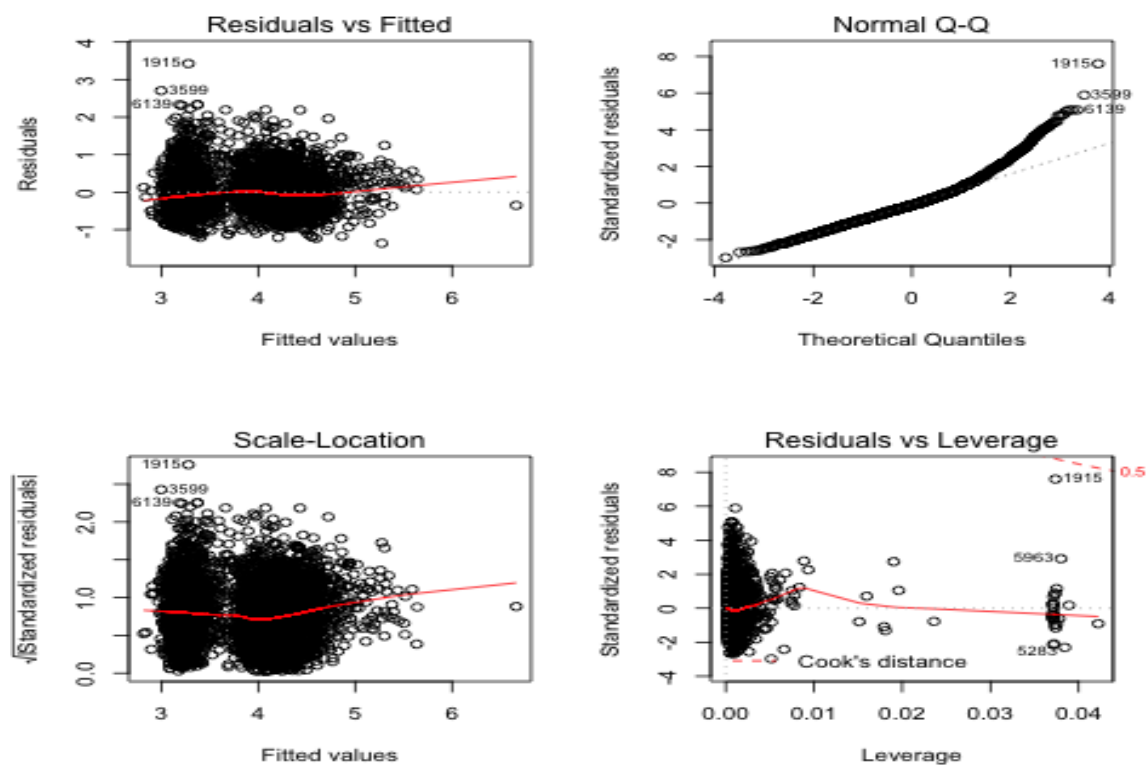


Figura A5 – Plot del modelo final (con atípicos)

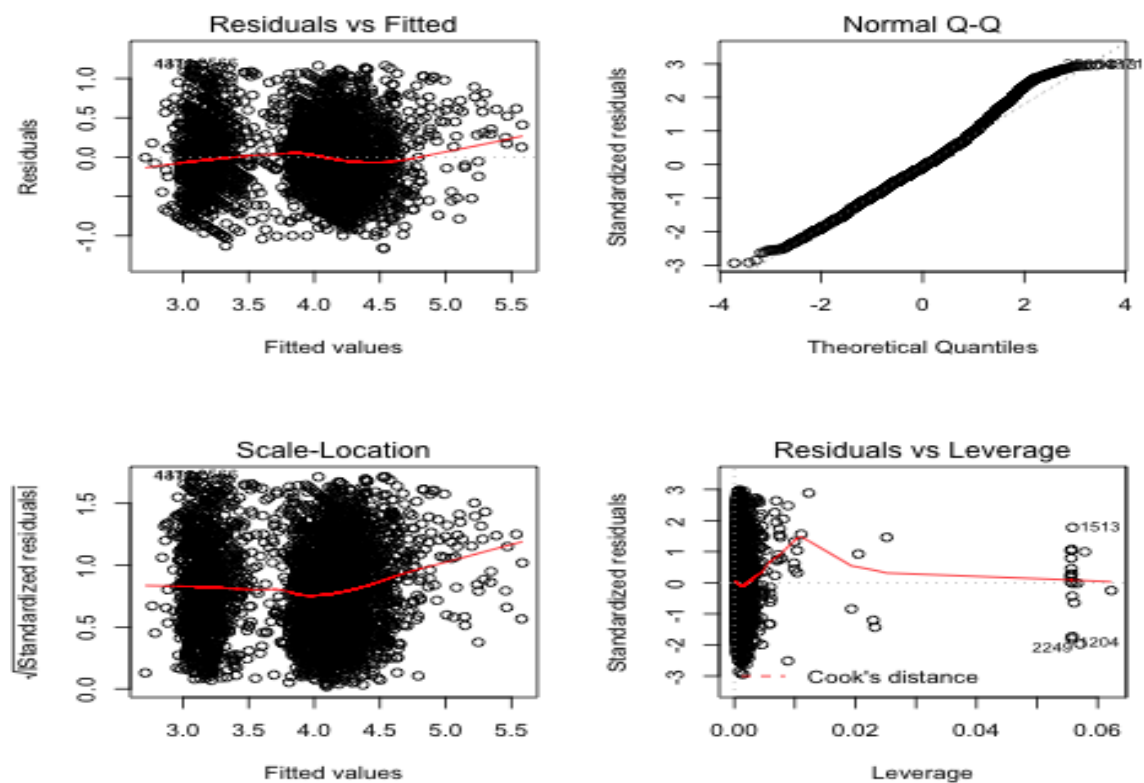


Figura A6 – Plot del modelo final (sin atípicos)

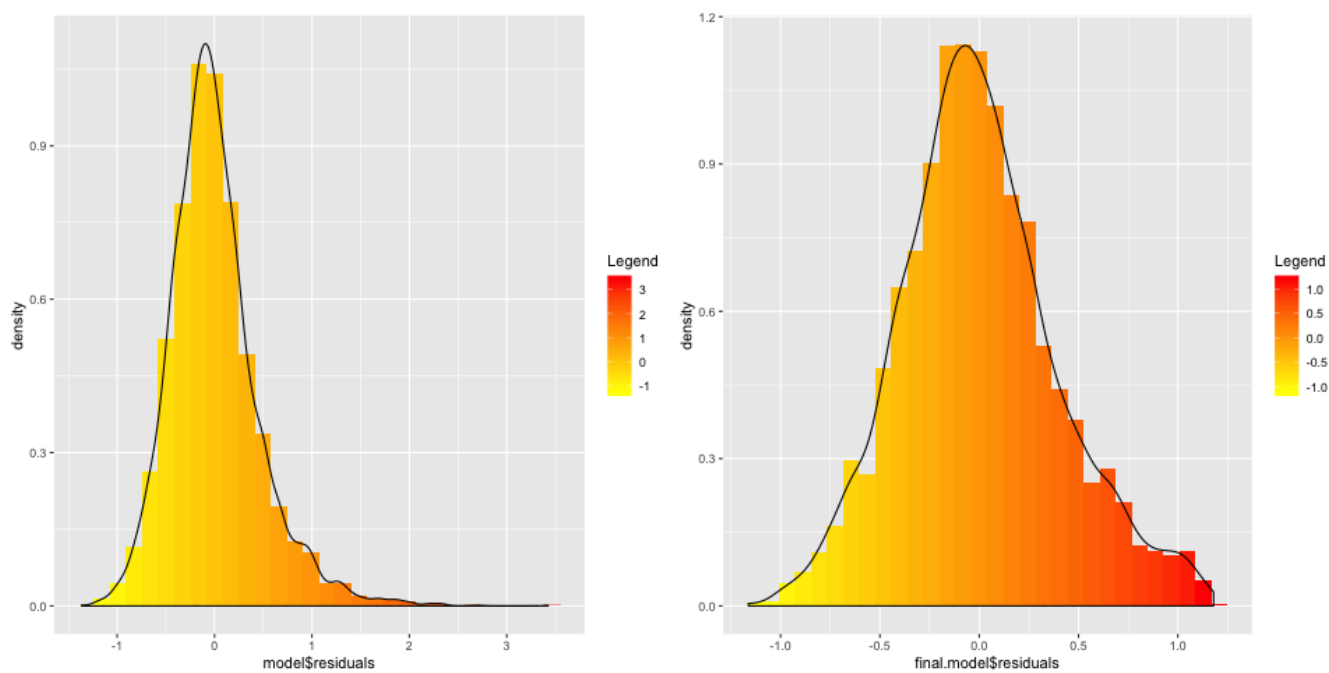


Figura A7 – Distribución de los residuos de los modelos con y sin valores atípicos

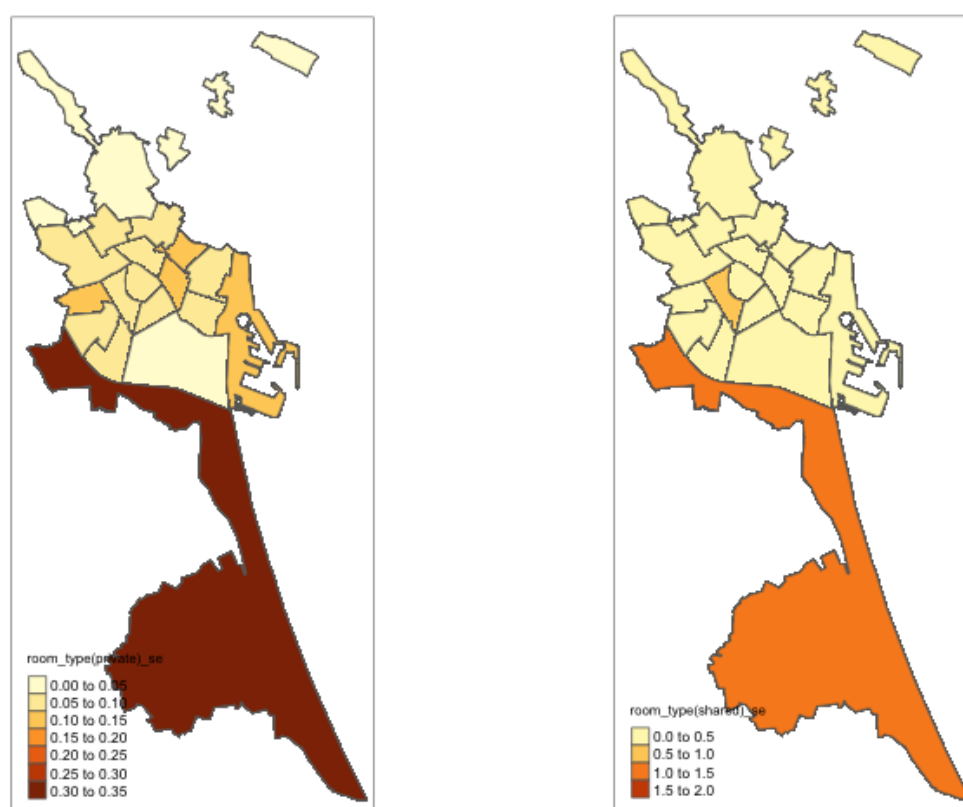


Figura A8 – Mapas de errores estándar de “room_type” por distritos

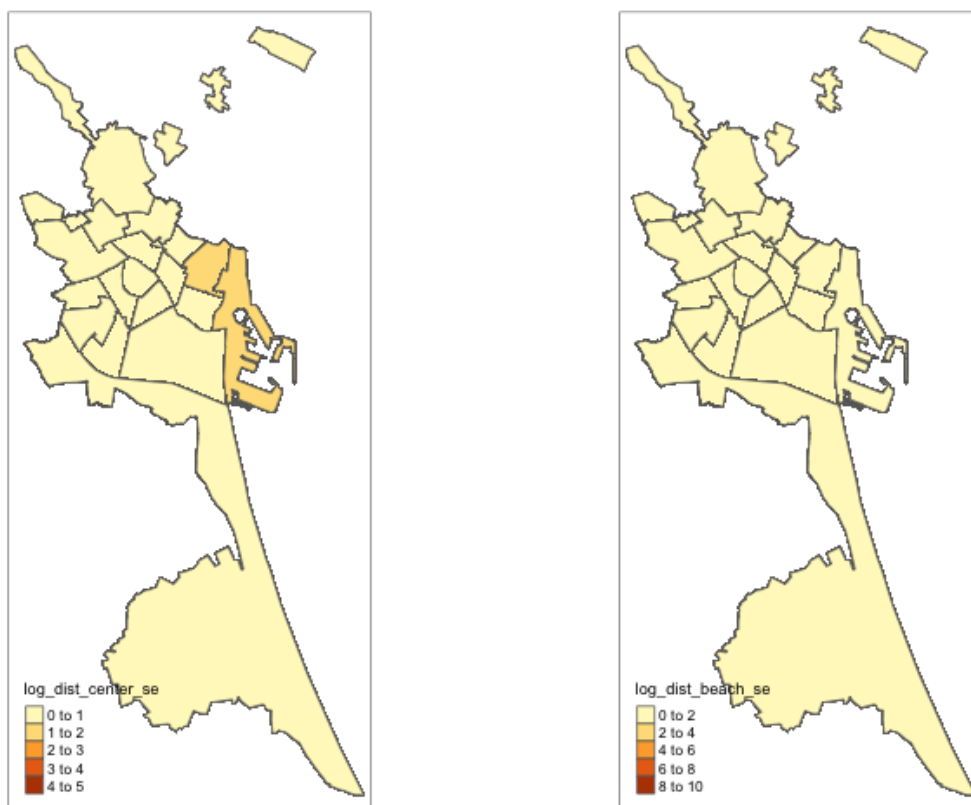


Figura A9 – Mapas de errores estándar de “log_distance_center” y “log_distance_beach” por distritos